

BANCOS DE DATOS NO BIBLIOGRAFICOS

La necesidad de información para el desarrollo de la ciencia y de la tecnología y en último extremo, para una más consciente toma de decisiones es, en el momento actual, universalmente reconocida. Pero «mantenerse informado» tanto a nivel individual como corporativo es cada vez más difícil debido al constante aumento de innovaciones científicas y técnicas y a la velocidad con que éstas se van produciendo, aunque se estén aplicando progresos técnicos para favorecer el acceso y la difusión de información, concretamente el uso de computadores y la organización y acceso a bancos de datos como servicio público o semipúblico de información a través de centros de documentación que actúan de intermediarios entre los productores y distribuidores de bancos o bases de datos y el público a que van dirigidos.

Los bancos de datos pueden ser clasificados de una manera general en bibliográficos y no bibliográficos, existiendo un número mayor de estos últimos. A pesar de ello, el especialista de la información o el centro de documentación como servicio público han estado más orientados hacia los bancos de datos bibliográficos. Existe una razón para explicar este punto. El uso de bancos de datos no bibliográficos requiere en la mayoría de los casos de ciertos conocimientos previos sobre la materia de que trata el banco de datos. Como consecuencia de ello, los productores de bancos de datos no bibliográficos han dirigido más sus actividades de marketing directamente hacia el usuario, sin intermediarios.

Antes de entrar más en el tema, quizá sea conveniente detenernos en uno de los puntos que más problemas ha ocasionado en la preparación de este trabajo, nos referimos al problema terminológico. El campo de la información científica carece, en términos generales, de una terminología claramente delimitada en español. Debido a ello comenzar definiendo conceptos tan elementales como «banco de datos» y «datos» no sea en modo alguno superfluo. Las diferentes clasificaciones de bancos de datos y sistemas que soportan estos bancos de datos serán tratadas a lo largo del presente artículo.

Martin define un banco de datos como «una colección de datos almacenados juntos con la menor redundancia posible para servir uno o más fines en una ordenación óptima». Encontrar una definición tan simple de «datos» en el campo informativo no resulta tan fácil. Una definición válida es la propuesta por Van Olphen, quien define «datos» como «fragmentos

de información real»; tales «fragmentos» pueden ser: a) tomados como referencias bibliográficas, en este caso los datos no ofrecen una información final, sino que son sólo intermediarios hacia la fuente final, y b) datos o valores finales sobre un cierto concepto, en este caso proporcionan acceso directo a la información. Este último aspecto es el que se refiere a los bancos de datos no bibliográficos y en español responde más al concepto asignado a la palabra «información» que al de «datos». En este punto, resulta conveniente precisar que los bancos a los que hace referencia este artículo son solamente aquellos orientados a producir información con fines comerciales. Por tanto, son excluidos aquellos de carácter interno como pueden ser los producidos a nivel empresa para elaborar nóminas, cotizaciones...

La nomenclatura aplicada a los bancos de datos «informativos» depende del enfoque que se le dé. Atendiendo al tipo de información que el banco de datos contiene o al sistema que lo soporta se pueden precisar dos diferentes clasificaciones. Según el primer aspecto, el tipo de información que el banco de datos contiene (Wanger y Landeu), éstos se dividen en:

A. Bancos de datos de referencia o secundarios, que actúan como intermediarios de información, ya que refieren a una determinada información o fuente primaria. Estos comprenden:

a.1. Bibliográficos, que son aquellas bases de datos que contienen citas bibliográficas y abstracts refiriendo a una fuente bibliográfica primaria.

a.2. De referencia, que contienen referencia a una información primaria, pero esta información no es bibliográfica como, por ejemplo, descripción de proyectos, instituciones, actividades...

B. Bases de datos finales o primarios que contienen información primaria. A su vez pueden dividirse en:

b.1. Bases de datos numéricos, que contienen información numérica. Esta información proviene más de la literatura técnica que directamente de investigación en laboratorio.

b.2. Bases de datos textual-numéricos, que combinan información numérica y textual.

b.3. Bases de datos de propiedades, que contienen principalmente valores asignados a componentes y reacciones físicas y químicas.

b.4. Bases de datos de texto completo. Aunque clasificadas en este grupo, ya que ofrecen una información final, están más cerca de las bases de datos bibliográficos que de las numéricas. Los problemas y posibilidades del sistema que los acompaña son básicamente los mismos y normalmente los mismos productores de bases de datos bibliográficos se dedican a la producción de bases de datos de texto completo.

Una segunda clasificación puede encontrarse atendiendo al sistema que soporta la base de datos. Igualmente encontramos dos grupos:

El primer grupo abarcaría sistemas de bancos de datos bibliográficos que incluiría, junto con los dos grupos propuestos en la división anterior para bancos de datos bibliográficos, el de bancos de datos de texto completo. Esta es la categoría que como se dijo al principio más atención ha recibido como campo comercial de «venta de información».

En segundo lugar encontramos los sistemas de bases de datos numéricos que incluye datos numéricos, textual-numéricos (aunque la clasificación de este último grupo en este apartado o en el primero es dudosa) y de propiedades. La principal característica de estos sistemas es que permiten la manipulación de los datos que incluyen.

Tratándose de bases de datos numéricos resulta extremadamente difícil separar el banco de datos del sistema que lo soporta y en la literatura sobre el tema, en la mayoría de los casos, ambos conceptos están mezclados. Las siguientes características son comunes para las bases de datos numéricos a los que hace referencia este artículo:

1. Son disponibles «on-line». Dado que una de las ventajas de estas bases de datos es que permiten la manipulación de más de un dato interactivamente, el uso de datos en «batch» se hace poco práctico.

2. Los productores de estas bases de datos no son, en principio, los usuarios a los que está dirigido el banco de datos, es decir, el fin principal es «vender información».

3. Estas bases de datos están disponibles para usuarios en todo el mundo a través de redes de telecomunicación.

Las compañías que ofrecen bancos de datos numéricos ofrecen estos servicios en tiempo compartido y generalmente están especializados en un solo tema, a diferencia de las compañías que ofrecen bases de datos bibliográficos que están orientados a varios temas. Hasta hace pocos años las organizaciones que ofrecían servicios en bancos de datos bibliográficos eran diferentes a las que ofrecían servicios en numéricos. Hoy en día esta diferenciación tiende a desaparecer, tal es el caso de Dialog, sistema producido por Lockheed, uno de los sistemas estadounidenses más extendido en todo el mundo y que incluye ambos servicios.

Los temas que abarcan las bases de datos numéricos son, por orden de importancia (en términos cuantitativos):

1. Economía; específicamente economía, estadística, cambios de moneda, información financiera sobre organizaciones o empresas específicas, marketing e información industrial. La información proviene tanto de fuentes gubernamentales como privadas (universidades u organizaciones).

2. Ciencia y tecnología. En este área en la mayoría de los casos el «software» viene junto a la base de datos, en otras palabras, el sistema que soporta la base de datos ha sido preparado específicamente para ella.

3. En último lugar, Ciencias Sociales. Todas las bases de datos en este área tienen posibilidades de operaciones matemáticas. Los datos provienen en su mayor parte de encuestas desarrolladas por destacadas organizaciones. Los censos nacionales pueden proporcionar importante información en el primer y tercer apartado.

La recolección y preparación de bancos de datos numéricos es mucho más cara que la de bancos de datos bibliográficos, además de que el proceso es más complejo, ya que requiere la evaluación de la información a incluir en la base de datos. Comprobar la veracidad de una información se complica aún más cuando aparecen diferentes valores para un mismo

parámetro. Hay diferentes soluciones a este problema; una fácil, pero cara y que complica la base de datos, es almacenar todos los valores encontrados para un mismo concepto. Si por el contrario se decide almacenar un solo valor, éste ha de ser decidido basándose en algún delimitador pre-establecido, bien sea la opinión de expertos o el elegir la fuente que presente mayor credibilidad.

Otro factor que da complejidad a estas bases de datos es que tienen que ser suficientemente flexibles como para que se pueda incluir todos los cambios ocurridos en los valores almacenados debidos al progreso de un determinado campo.

Se entiende por sistema de bases de datos numéricos «on-line» a la reunión de una o más bases de datos numéricos con un programa de recuperación de información capaz de presentar los datos «on-line». La presencia dentro del sistema de un programa que permita la manipulación de los valores almacenados es requisito casi imprescindible. Su falta dentro del sistema puede ser paliada previendo la compatibilidad del sistema con algún programa ya existente de operaciones matemáticas o estadísticas.

En comparación con los sistemas necesarios para operar bancos de datos bibliográficos, los sistemas para información numérica presentan un mayor costo de diseño, desarrollo y operación y su mercado es más pequeño, aunque no sus posibilidades inherentes.

Durante 1980 se ha estimado que se han desarrollado cuatro millones de búsquedas bibliográficas en USA y más de dos millones y medio en Europa. No hemos encontrado estimaciones semejantes para búsquedas o utilización de datos no bibliográficos; además de que el carácter, aún existente, de ser dirigido al usuario haría muy difícil esta estimación, pero las posibilidades técnicas actuales y su continuo avance hacen prever un fuerte incremento en el futuro, ya que técnicos y científicos tienen mayor necesidad de información primaria, de datos concretos, que de información bibliográfica.

Desde España son accesibles 210 bases de datos; de ellas, 84 son consideradas numéricas o factuales por la guía «Bases de Datos del Mundo», de las que la mitad son ofrecidas por I. P. Sharp Associates, que sólo se dedica a la producción de este tipo de bases de datos.

Los actuales desarrollos técnicos, especialmente la evolución del microcomputador y el continuo descenso en coste relativo de las telecomunicaciones (si bien enmascarado por la inflación), hacen que en el momento actual el presentar información primaria «on-line» sea técnica y económicamente posible. En cambio, creemos que otros problemas, principalmente de carácter político, psicológico y legal, quedan aún por superar para un definitivo uso de medios mecánicos para la transmisión de información primaria.

JOSEFINA ALJARO

BIBLIOGRAFIA

- Bases de Datos del Mundo. Catálogo de los sistemas de información científica, tecnológica, social y económica accesibles desde España por medios telemáticos. Madrid, Alhambra, 1981.
- BERNSTEINT, HERBERT, and LAWRENCE C. ANDREWS: «NIH-EPA Chemical Information System», *Database*, March, 1979, 35-49.
- BLUE, RICHARD I.: «Directory of U. S. Full-Text System Vendors», *On-Line Review*, vol. 3, núm. 2, June 1979, 175-179.
- BURCHINAL, LEE G.: «Notes from OSIS», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 7, 33.
- COMPTON, W. DALE: «Our Objective: Systems with Critically Evaluated Data», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 3, 37.
- FRIED, JOHN B.; JAMES A. LUEDKE, and STEPHEN A. RUBIN: «Online Numeric Database Systems», *On-Line Review*, vol. 1, núm. 3, July 1977, 70-72.
- HARRIS, CURT L., and SUZANNE H. ROBERTS: «The Search for Tomorrow. Low Cost, Full Text Search with Minimum Front End Investment». In: Brenner, Everett H., comp. *The Information Age in Perspective: Proceedings of the ASIS Annual Meeting*. New York, November 13-17, 1978. Volume 15. Abstracts of Papers. White Plains, N. Y., ASIS, 1978, 368-372.
- HELLER, S. R.; G. W. MILNE, and R. J. FELDMANN: «A Computer-Based Chemical System», *Science*, vol. 195, núm. 4275, January 1977, 253-259.
- HILLMAN, DONALD J.: «Model for the On-Line Management of Knowledge Transfer», *On-Line Review*, vol. 1, núm. 1, March 1977, 23-30.
- JARDINE, D. A.: «Data Base Data Structures». *Encyclopedia of Computer Science and Technology*. New York, Dekker, Inc., 1977, vol. 7, 86-114.
- KISSMAN, HENRY M.: «Building and On-Line Data Retrieval Systems», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 16, 36-37.
- LERNER, RITA G.: «Locating and Describing Data in Primary Journals», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 13, 35.
- LEVITAN, KAREN, and SHASHI SOOD: «A State-of-the-Art System: The Chemical Substances Information Network». In: Brenner, Everett H., comp. *The Information Age in Perspective: Proceedings of the ASIS Annual Meeting*. New York, November 13-17, 1978. Volume 15. Abstracts of Papers. White Plains, N. Y., ASIS, 1978, 202-205.
- LIDE, DAVID R.: «The National Standard Reference Data System», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 10, 34.
- LIDE, DAVID R.: «Quality Control of Data in the National Standard Reference Data System», In: Fry, Bernard M., comp. *Information Management in the 1980's: Proceedings of the ASIS Annual Meeting*. Chicago, September 26-October 1, 1977. Part One: Abstracts of Papers. White Plains, N. Y., ASIS, 1977, 117.
- LOEBL, ANDREW S.: «The Regional and Urban Studies Information Center (RUSTIC) of Oak Ridge National Laboratory». In: Fry, Bernard M., comp. *Information Management in the 1980's: Proceedings of the ASIS Annual Meeting*. Chicago, September 26-October 1, 1977. Part One: Abstracts of Papers. White Plains, N. Y., ASIS, 1977, 119-122.
- LUEDKE, JAMES A.: «Numeric Data Bases On-Line», *On-Line Review*, vol. 1, número 3, September 1977, 207-215.
- LUEDKE, JAMES A.; GABOR J. KOVACS, and JOHN B. FRIED: «Numeric Data Bases and Systems», *Annual Review of Information Science and Technology*, Volume 12, 1977, 119-181.
- MARRON, BEATRICE, and DENNIS FIFE: «Online Systems; Techniques and Services». *Annual Review of Information Science and Technology*, Volume 11, 1976, 163-210.

- MARTIN, JAMES: *Computer Data Base Organization*. Englewood Cliffs, New Jersey, Prentice Hall, Inc., 1975.
- «Nonbibliographic Databases Online», *On-Line Review*, vol. 3, núm. 2, June 1979, 153-174.
- Predicasts Terminal System*. Users Manual. Cleveland, Ohio, Predicasts, Inc., 1979.
- RINEWALT, J. SR.: «Feature evaluation of a Full-Text Information-Retrieval System», *On-Line Review*, vol. 1, núm. 1, March 1977, 43-52.
- ROWE, JUDITH S., and MARY RYAN: «Library Service from Numerical Data Bases: The 1970 Census as a Paradigm», *College & Research Libraries*, vol. 35, January 1974, 7-17.
- SHUMAN, JACK NORMAN: «The Data Revolution: Its Policy Implications», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 11-12, 34-35.
- SIMON, RICHARD C.: «Combined Online Numeric Text Analysis and Retrieval Systems», In: Martin, Susan K. comp. *Information-Politics: Proceedings of the ASIS Annual Meeting*. San Francisco, October 4-9, 1976. Part One: Abstracts of Papers. Volume 13. Washington, ASIS, 1976, 71.
- SIMON, RICHARD C., and TAMAS E. DOSZKOCs: «Implementation of a User Interface for a Complex Numeric Textual Data Base». In: Brenner, Everett H. comp. *The Information Age in Perspective: Proceedings of the ASIS Annual Meeting*, New York, November 13-17, 1978. Volume 15. Abstracts of Papers. White Plains, N. Y. ASIS, 1978, 309-313.
- TOULOUKIAN, Y. S.: «A Profile of the Center for Information and Numerical Data Analysis and Synthesis (CINDAS)». In: Fry, Bernard M., comp. *Information Management in the 1980's: Proceedings of the ASIS Annual Meeting*. Chicago, September 26-October 1, 1977. Part One: Abstracts of Papers. White Plains, N. Y., ASIS, 1977, 114.
- VAN OLPHEN, HENDRIK: «The Numerical Data Advisory Board», *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 8, 9, 33.
- VAN OLPHEN, HENDRIK: «What Are» Data?, *Bulletin of the ASIS*, vol. 1, núm. 7, February 1975, 8, 9.
- WANGER, JUDITH, and RUTH N. LANDAU: «Nonbibliographic On-Line Data Base Services», *Journal of the ASIS*, May 1980, 171-180.
- WILLIAMS, MARTHA E.: «Online Retrieval Today and Tomorrow», *On-Line Review*, vol. 2, núm. 4, December 1978, 353-366.